Crystal Structure of NusA from *Thermotoga Maritima* and Functional Implication of the N-Terminal Domain†

Dong Hae Shin,‡ Henry Huy Nguyen,‡,§ Jaru Jancarik,|| Hisao Yokota,‡ Rosalind Kim,‡ and Sung-Hou Kim*,‡,||

Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, and Departments of Molecular and Cell Biology and Chemistry, University of California, Berkeley, California 94720

Received June 30, 2003; Revised Manuscript Received September 19, 2003

ABSTRACT: We report the crystal structure of N-utilizing substance A protein (NusA) from *Thermotoga maritima* (*Tm*NusA), a protein involved in transcriptional pausing, termination, and antitermination. *Tm*NusA has an elongated rod-shaped structure consisting of an N-terminal domain (NTD, residues 1−132) and three RNA binding domains (RBD). The NTD consists of two subdomains, the globular head and the helical body domains, that comprise a unique three-dimensional structure that may be important for interacting with RNA polymerase. The globular head domain possesses a high content of negatively charged residues that may interact with the positively charged flaplike domain of RNA polymerase. The helical body domain is composed of a three-helix bundle that forms a hydrophobic core with the aid of two neighboring β-strands. This domain shows structural similarity with one of the helical domains of σ^{70} factor from *Escherichia coli*. One side of the molecular surface shows positive electrostatic potential suitable for nonspecific RNA interaction. The RBD is composed of one S1 domain and two K-homology (KH) domains forming an elongated RNA binding surface. Structural comparison between *Tm*NusA and *Mycobacterium tuberculosis* NusA reveals a possible hinge motion between NTD and RBD. In addition, a functional implication of the NTD in its interaction with RNA polymerase is discussed.

DNA-dependent RNA polymerases utilize several protein cofactors to facilitate transcription. These proteins serve to guide transcription by recognizing signal sequences along the DNA template/RNA and thus control a host of transcriptional processes. N-utilizing substance A protein (NusA)[1] is essential in modulating elongation through association with the core component of RNA polymerase (RNAP) after the σ^{70} initiation factor is released from a promoter (*1*). NusA influences elongation by increasing the dwell time for RNAP at certain pause sites (*2*−*4*), possibly by interacting with and stabilizing the RNA hairpin structure associated with pause sites (*5*, *6*).

Sequence and structural alignments have suggested that NusA has both S1 and KH homology domains that are known to bind to RNA (*7*). The S1 and K-homology (KH) domains of NusA are important for enhancing both transcriptional termination and antitermination. The *Escherichia coli* NusA possesses two RNA polymerase-binding domains, one in the amino-terminal 137 amino acids and the other in the carboxy-terminal 264 amino acids (*7*). Amino-terminal RNA polymerase-binding domain provides a functional contact that enhances termination at an intrinsic terminator or antitermination by phage λ N protein. The carboxy-terminal domain is known to form an N−NusA−*nut* site or N−NusA−RNA polymerase−*nut* site complex. The instability of the complexes lacking this carboxy-terminal domain of NusA can be compensated for by the presence of additional *E. coli* elongation factors (*7*). NusA is the largest essential Nus protein ($M_r \sim 55\,000$) in *E. coli*. The NusA gene is also present in the *Mycoplasma genitalium* genome, the presumed minimal set of genes required for a living organism (*8*).

The crystal structures of NusA from *Thermotoga maritima* (*Tm*NusA-1) (*9*) and *Mycobacterium tuberculosis* (*Mt*NusA) (*10*) have been reported. They describe the role of RNA binding domain in detail. However, NTD was ambiguously determined in the crystal structure of *Tm*NusA-1, and NTD of one molecule out of two in the asymmetric unit was invisible in the case of *Mt*NusA. Here, we report a completely ordered 2.5 Å structure of NusA from *T. maritima* (*Tm*NusA-2) and discuss the possible role of its NTD in the interaction with RNA polymerase.

EXPERIMENTAL PROCEDURES

Cloning. Primers (Bioneer Corp., South Korea) for PCR amplification from genomic DNA contained an *Nde*I restriction site in the forward primer (5′-CATATGAACATAG-

† The work described here was supported by the National Institutes of Health GM 62412.

* To whom correspondence should be addressed: e-mail SHKim@cchem.berkeley.edu; fax 1-510-486-5272.

‡ Lawrence Berkeley National Laboratory.

§ Department of Molecular and Cell Biology, UC Berkeley.

|| Department of Chemistry, UC Berkeley.

[1] Abbreviations: NusA, N-utilizing substance A protein; *Tm*NusA, N-utilizing substance A protein from *Thermotoga maritima*; NTD, N-terminal domain; RBD, RNA binding domains; KH, K-homology; RNAP, RNA polymerase; *Tm*NusA-1, N-utilizing substance A protein from *Thermotoga maritima* studied by Worbs et al. (see text); *Mt*NusA, *Mycobacterium tuberculosis* N-utilizing substance A protein; *Tm*NusA-2, N-utilizing substance A protein from *Thermotoga maritima* in this study; rms, root-mean-square; EDTA, ethylenediaminetetraacetic acid; Hepes, *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid; PCR, polymerase chain reaction; PEG, poly(ethylene glycol); Tris, tris-(hydroxymethyl)aminomethane.

10.1021/bi035118h CCC: $25.00
Published on Web 10/28/2003

GCTTGCTGGAAGC) and a *Bam*HI site in the reverse primer (5′-AGATCTTTACAGGTTCATGATCGGTTTTAT-GTC). PCR was performed with Deep Vent Polymerase (New England Biolabs, Inc., Beverly, MA) and genomic DNA (American Type Culture Collection, Manassas, VA). The PCR product was cloned into pCR-BluntII-TOPO vector (Invitrogen Corp., Carlsbad, CA) and the *Tm*NusA gene (gi_4982355) insert was confirmed by DNA sequencing. The amplified TOPO vector was restricted with *Nde*I and *Bam*HI and the gene insert was purified by agarose gel electrophoresis extraction. The insert was ligated into pSKB3 (gift from Steve Burley, Structural Genomics, San Diego, CA), digested with *Nde*I and *Bam*HI, and transformed into DH5α. The expression plasmid construct was transformed into Rosetta.pLysS.RARE *E. coli* cells (Novagen, Inc., Madison, WI).

Bacterial Expression and Protein Purification. The Rosetta·pLysS.RARE *E. coli* cells harboring pSKB3/$(His)_6$-*Tm*NusA plasmid were grown at 37 °C in M9 medium containing 50 μg/mL kanamycin and 34 μg/mL chloramphenicol until they reached an OD_{600} of 0.8. The incubating temperature was then lowered to 18 °C and the selenomethionine (Se-Met) *Tm*NusA protein was prepared according to the method of Doublie (*11*). The Se-Met *Tm*NusA protein was induced with 0.3 mM isopropyl β-D-thiogalactopyranoside (IPTG) for 3 h. The cells were harvested by centrifugation (5000 *g* for 10 min) and stored at −80 °C. The cell pellet was resuspended in 50 mM Tris, pH 7.5, frozen in liquid nitrogen, and thawed on ice. Protease inhibitors consisting of phenylmethanesulfonyl fluoride (PMSF), leupeptin, pepstatin, antipain, and chymostatin were added to final concentrations of 1 mM, 10 μg/mL, 2.5 μg/mL, 1 μg/mL, and 5 μg/mL, respectively. DNase I was added at a final concentration of 1 mg/mL. The cells were sonicated and spun at 12000*g* to remove cell debris. The cleared lysate was incubated at 80 °C for 20 min to precipitate heat-unstable proteins from *E. coli*. The supernatant was collected after centrifugation for 30 min at 40 000 *g* in a Beckman Ti70 rotor, filtered through a 0.2 μm filter, and brought to 0.3 M NaCl. The clarified supernatant was mixed with Talon IMAC resin (BD Biosciences Clontech, Palo Alto, CA) previously equilibrated with buffer A (50 mM Tris-HCl and 0.3 M NaCl, pH 7.5) for 1 h on a tilt shaker in a batch method. The flowthrough was collected after centrifugation at 700*g* for 5 min. The resin was washed twice with 10 bed volumes of buffer A. The $(His)_6$-*Tm*NusA protein was then eluted with 300 mM imidazole. Fractions containing $(His)_6$-*Tm*NusA protein were pooled and diluted 10-fold before they were applied on a 5-mL HiTrap Q column (Amersham Biosciences Corp., Piscataway, NJ) previously equilibrated with buffer B (20 mM Tris-HCl and 1 mM EDTA, pH 7.5). The column was washed with 6 column volumes (CV) of buffer B. A 20 CV linear gradient from 0 to 1 M NaCl in buffer B was applied. Fractions containing $(His)_6$-*Tm*NusA protein were pooled and concentrated by using a 10K Ultrafree unit (Millipore, Billerica, MA) and finally dialyzed against buffer C (20 mM Tris-HCl, 1 mM EDTA, and 0.1 M NaCl, 5% glycerol, pH 7.5). The protein concentration measured by absorbance at 280 nm was calculated by use of an extinction coefficient of 27 880 $M^{-1}cm^{-1}$. Since we did not cleave the $(His)_6$ tag, the 25-residue tag (MGSSHHHHHHDYDIPTTENLYFQGH) remained fused to the N-terminus of the protein.

Table 1: Statistics of X-ray Diffraction Data and Structure Refinement[a]

(A) Comparison of Crystal Parameters and Refinement Statistics

	*Tm*NusA-2	*Tm*NusA-1
wavelength (Å)	0.953 72	0.950 00
space group	$P4_32_12$	$P4_32_12$
cell dimension $a = b$ (Å)	116.2	115.5
cell dimension c (Å)	64.6	63.8
volume fraction of protein (%)	52.1[a]	54.7
V_m ($Å^3$/Da)	2.67	2.82
total no. of residues	344	344
total non-H atoms	2667	2667
no. of water molecules	228	375
average temperature factors ($Å^2$)		
protein	59.9	66.4
solvent	62.4	68.4
resolution range of reflections used (Å)	20.0−2.5	20.0−2.1
amplitude cutoff	0.0	0.0
R factor (%)	22.3	24.4
free R factor (%)	29.8	31.9
stereochemical ideality: bond		
bond (Å)	0.011	0.011
angle (deg)	1.58	1.66

(B) Statistices of the Three Wavelength MAD Data Sets

data set	edge	peak	remote
wavelength (Å)	0.979 23	0.979 04	0.953 72
resolution (Å)	43.0−2.5	43.0−2.5	43.0−2.5
redundancy	4.8(3.1)[b]	10.5 (8.3)	4.9(3.1)
unique reflections	29 460 (1326)	29 536 (1408)	29 503 (1382)
completeness (%)	99.4 (90.1)	99.8 (96.8)	99.7 (95.0)
I/σ	23.8 (1.6)	34.1 (3.9)	24.3 (2.3)
R_{sym}[c] (%)	6.6 (74.5)	6.9 (43.4)	6.4 (52.8)

[a] The 25-residue $(His)_6$ tag was included to calculate a V_m value, though it was not visible in the electron density map. [b] Numbers in parentheses refer to the highest resolution shell, which is 2.50−2.54 Å for all wavelength data. [c] $R_{sym} = \Sigma_{hkl}\Sigma_i|I_{hkl,i} - \langle I\rangle_{hkl}|/\Sigma|\langle I\rangle_{hkl}|$.

The initial crystallization conditions were screened by the sparse matrix method (*12*) with a Hampton Research kit (Laguna Niguel, CA) at room temperature. One microliter of 20 mg/mL protein solution in 20 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.2 M NaCl, and 5% glycerol was mixed with one microliter of reservoir solution of 2% PEG 400, 2.0 M ammonium sulfate, and 0.1 M Hepes, pH 7.5. The hanging drop was equilibrated with 0.5 mL of reservoir solution. Ellipsoidal-shaped crystals grew in a week to approximate dimensions of 0.6 mm × 0.2 mm × 0.2 mm.

Data Collection and Reduction. The crystals were soaked in a drop of mother liquor with 16% glycerol (about 10 μL) for about 1 min before being flash-frozen in liquid nitrogen and exposed to X-ray. X-ray diffraction data sets were collected at three wavelengths at the Macromolecular Crystallography Facility, beamline 5.0.2, at the Advanced Light Source at Lawrence Berkeley National Laboratory by use of an Area Detector System Co. (Poway, CA) Quantum 4 CCD detector placed 140 mm from the sample. The oscillation range per image was 1.0° with no overlap between two contiguous images. X-ray diffraction data were processed and scaled with DENZO and SCALEPACK from the HKL program suite (*13*). The synchrotron data were collected to 2.5 Å. Data statistics are summarized in Table 1B. The crystal belongs to the primitive tetragonal space group $P4_32_12$, with unit-cell parameters of $a = b = 116.2$ Å and $c = 64.6$ Å.

Structure Determination and Refinement. Both the multiwavelength anomalous dispersion (MAD) and single-

wavelength anomalous dispersion (SAD) experiments were performed. However, we only used peak data set, because the redundant SAD experimental electron density map was of better quality. The program SOLVE (*14*) was used to locate the selenium sites in the crystal and the program SHARP (*15*) was used for phase refinement to 2.8 Å followed by solvent flattening with SOLOMON (*16*). The initial SAD phases were good enough to trace the backbone structure of the protein. The model building was performed with O (*17*). A model containing all 344 *Tm*NusA-2 residues was derived from progressive improvement of the electron density map by rounds of phase combination and manual building. However, none of the 25-residue $(His)_6$ tag was modeled due to the lack of electron density.

The program CNS was used for all refinement calculations (*18*). All the reflections in the peak data set between 20.0 and 2.5 Å were included throughout the refinement calculations. Ten percent of the data was randomly chosen for free *R*-factor cross validation. The refinement statistics are shown in Table 1A. Isotropic *B*-factors for individual atoms were initially fixed to 15 Å^2 and were refined in the last stages. The $2F_o - F_c$ and $F_o - F_c$ maps were used for manual rebuilding between refinement cycles and for the location of solvent molecules. Atomic coordinates have been deposited in the Protein Data Bank under the access code 1L2F.

RESULTS AND DISCUSSION

Quality of the Model. All residues are well-defined by the electron density for the refined model of *Tm*NusA-2 (Figure 1). The final model has been refined to 2.5 Å resolution with a crystallographic *R*-factor of 22.3%. The averaged *B*-factors for the main-chain and the side-chain atoms are 56.9 and 62.7 Å^2, respectively. In the model of *Tm*NusA-2, the residues showing the higher *B*-factors are located around the N-terminal loop (residues 1−4, 109.8 Å^2), the region around α-helix H3 (residues 78−88, 109.4 Å^2), and the C-terminal loop (residues 342−344, 119.8 Å^2). All residues lie in the allowed region of the Ramachandran plot produced with PROCHECK (*19*). Table 1A summarizes the refinement statistics as well as model quality parameters. The mean positional error in the atomic coordinates for the refined model is estimated to be within 0.33 Å by the Luzzati plot (*20*).

*Structural Features of TmNusA-2. Tm*NusA-2 is a highly elongated molecule with dimensions of 125 Å × 28 Å × 25 Å, due to five domains arranged in a linear manner (Figure 2). The NTD is divided into two subdomains, the globular head domain and the helical body domain that is mainly composed of three helices (H1, H2, and H4) (Figures 2 and 3). The NTD is followed by three RBDs: S1 domain, named after its identification in ribosomal protein S1 (21), folds into a five-stranded antiparallel β-barrel with Greek key topology and a small 3_{10}-helix following the third strand β7. Two KH domains are composed of mixed α/β. These two well-known RNA binding motifs, S1 and KH domains, form an elongated RNA binding surface, showing the first structural example of two connected different RNA binding domains as described in details by Worbs et al. (*9*).

Since there was a "two NusA" model proposed on the basis of functional biochemical studies of NusA (*22*), we examined molecular contacts in the crystal and performed size-exclusion chromatography (data not shown) to research the possibility of dimeric interaction. However, *Tm*NusA-2 indicated a clear monomeric form in solution and in the crystal without ambiguity.

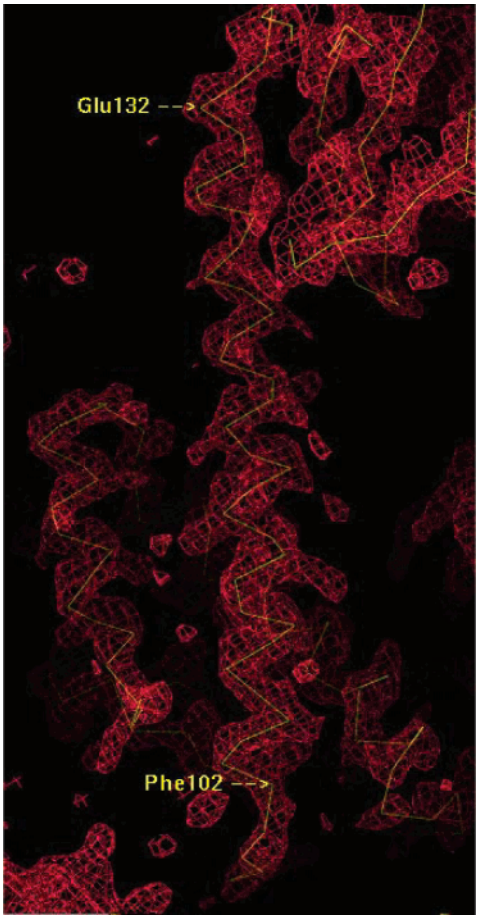

FIGURE 1: Initial electron-density map contoured at 1σ. The initial map after solvent flattening with program SOLOMON (*16*) was calculated from all reflection data between 20 and 3.2 Å. The N- and C-terminal residues of α-helix H4 are labeled. The red net represents the electron-density map. The yellow tracing represents a polyalanine model of *Tm*NusA-2.

Structural Comparison between TmNusA-1 and TmNusA-2. The secondary structure comparison between the two *Tm*NusAs is depicted in Figure 3, and the comparison of refinement statistics is listed in Table 1A. The root-mean-square (rms) differences between the two structures are 2.38 Å for all 344 Cα atoms, 2.48 Å for all main-chain atoms, and 3.19 Å for all 2260 atoms. The major structural differences are observed around NTD (residues 1−132) (Figure 4a). The rms differences between NTDs are 3.38 Å for 132 Cα atoms, 3.27 Å for the main-chain atoms, and 4.34 Å for all 1053 atoms. The structural feature of NTD (residues 1−132) can be divided into two subdomains: (a) the globular head domain and (b) the helical body domain. The globular head domain has a globular shape with minor structural differences between the two *Tm*NusA structures [e.g., the presence of β3 in *Tm*NusA-2 (Figure 3)]. The helical body domain forms an elongated hydrophobic core surrounded by three α-helices and two neighboring β-strands in the case of *Tm*NusA-2. This domain shows the largest structural difference between the two *Tm*NusA structures as shown in Figure 5 (residues 1−19, 6.41 Å for main-chain atoms; residues 98−108, 4.30 Å; residues 123−132, 5.27

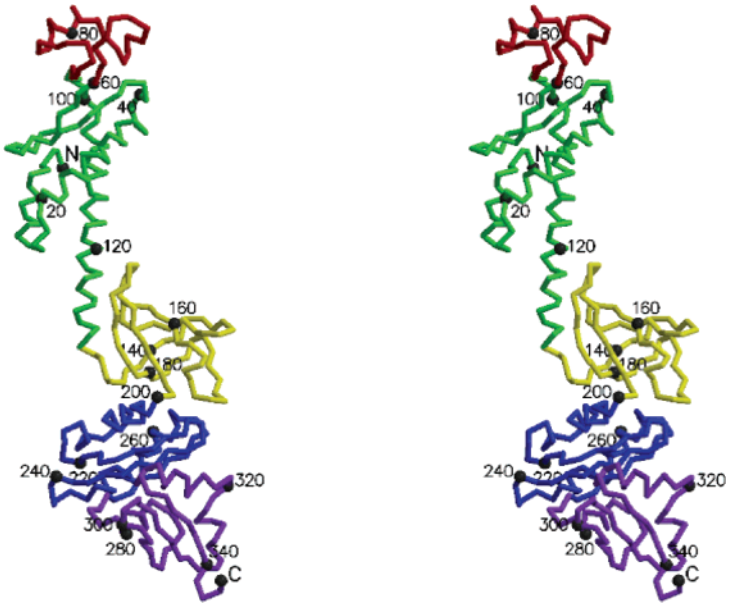

FIGURE 2: Stereo drawing of a Cα trace of TmNusA-2. The helical body domain (green), the globular head domain (red), S1 (yellow), KH1 (blue), and KH2 (purple) domains of *Tm*NusA-2 are represented as a thick line. Every twentieth residue is numbered and represented by a dot. The N- and C-termini and the secondary structural elements are labeled. The figure was generated by MOLSCRIPT (*31*).

```
                 H1          H2                 β1      β2          β3 H3      β4          
                 1  10        30        50        70        90       100 110
TmNusA-2         MNIGLLEALDQLEEEKGISKEEVIPILEKALVSAYRKNFGNSKNVEVVIDRNTGNIKVYQLLEVVEEVEDPATQISLEEAKKIDPLAEVGSIVKKELNVKNFGRIAAQTA 110
TmNusA-1         MNIGLLEALDQLEEEKGISKEEVIPILEKALVSAYRKNFGNSKNVEVVIDRNTGNIKVYQLLEVVEEVEDPATQISLEEAKKIDPLAEVGSIVKKELNVKNFGRIAAQTA 110
MtNusA           MNI-DMAALHAIEVDRGISVNELLETIKSALLTAYRHTQGHQTDARIEIDRKTGVVRVIARETDEAGN--------------------LISEWDDTPEGFGRIAATTA  85
                 ***  . **  .* ..***  *..  .  **..***.. *.     . *** ** ..*           *       ****** **

                 H4          β5      β6      β7  G1      β8      β9      H5
                 111 120 130 140 150 160 170 180 190 200 210
TmNusA-2         KQVLIQRIRELEKEKQFEKYSELKGTVTTAEVIRVM----GEWADIRIG----KLETRLPKKEWIPGEEIKAGDLVKVYIIDVVKTTKGPKILVSRRVPEFVIGLMKLEI 212
TmNusA-1         KQVLIQRIRELEKEKQFEKYSELKGTVTTAEVIRVM----GEWADIRIG----KLETRLPKKEWIPGEEIKAGDLVKVYIIDVVKTTKGPKILVSRRVPEFVIGLMKLEI 212
MtNusA           RQVMLQRFRDAENERTYGEFSTREGEIVAGVIQRDSRANARGLVVVRIGTETKASEGVIPAAEQVPGESYEHGNRLRCYVVGVTRGAREPLITLSRTHPNLVRKLFSLEV 197
                 .**..** *. * *. . .*   * . . . *           .***      *  .*  * .***    *  .. *.. * . .. * * .**  *  *  *  **.

                 β10     β11     H6      H7      β12     H8      β13     β14 G2  H9
                 213 220 230 240 250 260 270 280 290 300 310 320
TmNusA-2         PEVENGIVEIKAIAREPGVRTKVAVASNDPNVDPIGACIGEGGSRIAAILKELKGEKLDVLKWSDDPKQLIANALAPATVIEVEILDKENKAARVLVPPTQLSLAIGKGG 322
TmNusA-1         PEVENGIVEIKAIAREPGVRTKVAVASNDPNVDPIGACIGEGGSRIAAILKELKGEKLDVLKWSDDPKQLIANALAPATVIEVEILDKENKAARVLVPPTQLSLAIGKGG 322
MtNusA           PEIADGSVEIVAVAREAGHRSKIAVRSNVAGLNAKGACIGPMGQRVRNVMSELSGEKIDIIDYDDDPARFVANALSPAKVVSVSVIDQTARAARVVVPDFQLSLAIGKEG 307
                 **.  * *** *.*** * *.*.** **  ..    *****  * *.  .. ** ***.*..   *** . .****.** *. * ..*.  .****.**  ******** *

                 H10     β15
                 323 330 340                                    identity  homology
TmNusA-2         QNARLAAKLTGWKIDIKPIMNL------------------  344
TmNusA-1         QNARLAAKLTGWKIDIKPIMNL------------------  344
MtNusA           QNARLAARLTGWRIDIRGDAPPPPPGQPEPGVSRGMAHDR  347  38.4%    56.1%
                 *******.****.***.
```

FIGURE 3: Sequence comparison of *Tm*NusA and *Mt*NusA based on the crystal structures. The secondary structure derived from *Tm*NusA-2 is shown above the sequence. Blue dotted regions represent the sequences belonging to α-helices, green dots for 3_{10}-helix, and red dotted regions for β-strands. Most of the secondary structural elements are conserved between *Tm*NusA-2 and *Mt*NusA, except for β3 and H3 present only in *Tm*NusA-2 and one helix between β5 and β6 that is present only in *Mt*NusA. The shaded regions refer to different structural motifs and are maintained in Figure 2. A dash represents a gap in the sequence; asterisks are shown below conserved residues, and dots are shown below homologous residues. The sequence number refers to that of *Thermotoga maritima*.

Å). In *Tm*NusA-1, the N-terminal region (from 1 to 16) forms a loop structure with two hydrophilic residues (E7 and Q11) interspersed and disrupting the formation of a regular hydrophobic core. In the *Tm*NusA-2 structure, however, the N-terminal region forms an α-helix (H1) with two hydrophilic residues (E7 and Q11) protruding outward without disrupting the hydrophobic core. The middle helix (H4) connecting NTD to RBD is a 9-turn helix in *Tm*NusA-2 rather than a 4-turn helix plus two loops as shown in *Tm*NusA-1 (Figure 4a). With the aid of these two helices (H1 and H4), the NTD of *Tm*NusA-2 comprises a helical bundle different from that of *Tm*NusA-1.

The structural features described above for *Tm*NusA-2 are more consistent with other observations: First, the higher helical content in the NTD is strongly supported by a biochemical experiment by Mah et al. (*7*). They showed that the expressed NTD of *E. coli* NusA (residues from 1 to 137) has a helical content around 39% by CD analysis. Considering that the sequence homology in the NTD is around 57% (33% identity) between *E. coli* and *Thermotoga maritima*, this high ratio of helical content (42%) in *Tm*NusA-2 looks more plausible than the low helical content in *Tm*NusA-1 (25%). Second, the presence of the stable hydrophobic core in NTD is more consistent with the fact that *Tm*NusA-2 showed an extreme thermostability at 80 °C for 20 min during purification. The helical bundle of *Tm*NusA-2 forms a hydrophobic core with two neighboring β-strands, while *Tm*NusA-1 shows several charged side chains interfering with the formation of this core. In addition, the residues containing hydrophobic aromatic rings located on a loop of *Tm*NusA-1

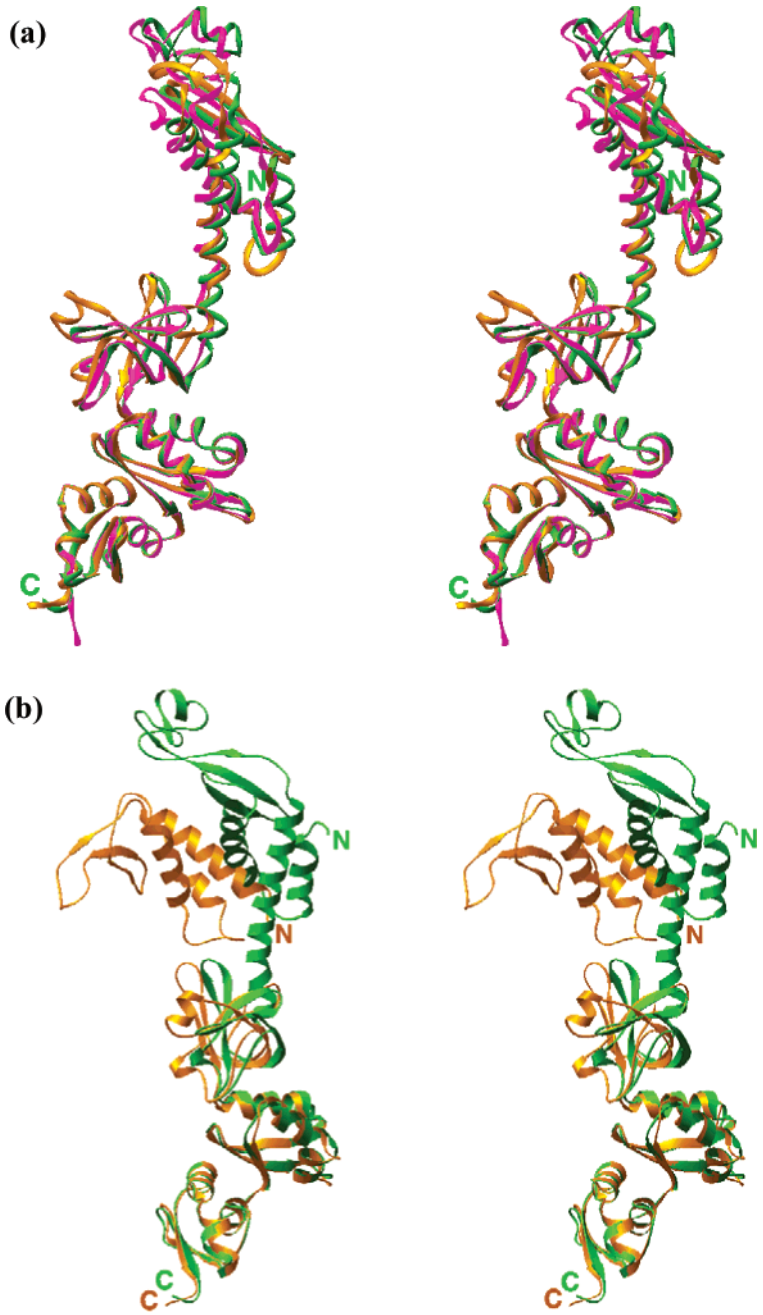

FIGURE 4: Structural comparison of NusAs. (a) Structural comparison of three NusAs after Cα rms fitting. Green represents *Tm*NusA-2, magenta for *Tm*NusA-1, and yellow for *Mt*NusA. The alignment of *Tm*NusA-2 and *Tm*NusA-1 is based on the entire Cα rms fitting. The alignment of *Tm*NusA-2 and *Mt*Nusa is based on Cα rms fitting of NTD and RBD, separately (see text). (b) Structural comparison of *Tm*NusA and *Mt*NusA. *Tm*NusA-2 structure is in green and *Mt*NusA structure in yellow. The alignment of *Tm*NusA-2 and *Mt*NusA is based on Cα rms fitting of two KH domains only (see text). The program RIBBONS (*32*) was used to generate the figures.

NTD (Phe39 and Phe102) protrude toward the solution rather than forming a hydrophobic core as shown in *Tm*NusA-2 (Figure 5a). The displacements of Cζ atoms of Phe39 and Phe102 are 10.33 Å (4.72 Å for Cα) and 15.25 Å (9.59 Å for Cα), respectively.

Actually, the crystallization conditions of the two structures were almost the same except for two things: (1) *Tm*NusA-2 had an N-terminal $(His)_6$ tag of 25 residues and (2) *Tm*NusA-2 had been crystallized in 2.0 M ammonium sulfate (AMS), which was higher than that for *Tm*NusA-1 (1.8 M AMS). Therefore, for the identification of the exact source of difference in conformational change of NTD, more experiments should be done. However, we did not see any visible electron density of a $(His)_6$ tag and thus postulate that the $(His)_6$ tag might not influence the conformational change of the NTD.

Since Worbs et al. (*9*) mentioned the ambiguous structural determination in NTD due to poor electron density quality as reflected by both high *R*-factor and high free *R*-factor (Table 1A), it is more plausible that some region of the NTD could have been wrongly determined in *Tm*NusA-1. However, we cannot totally exclude the possibility of a potential structural flexibility or a conformational transition as an explanation of regions of unstructured *Tm*NusA-1 NTD.

The structure of the three RNA binding domains in the rest of the molecule is almost the same in both structures. The rms differences of this region between the two structures (from 133 to 342) are 0.53 Å for 210 Cα atoms, 0.67 Å for

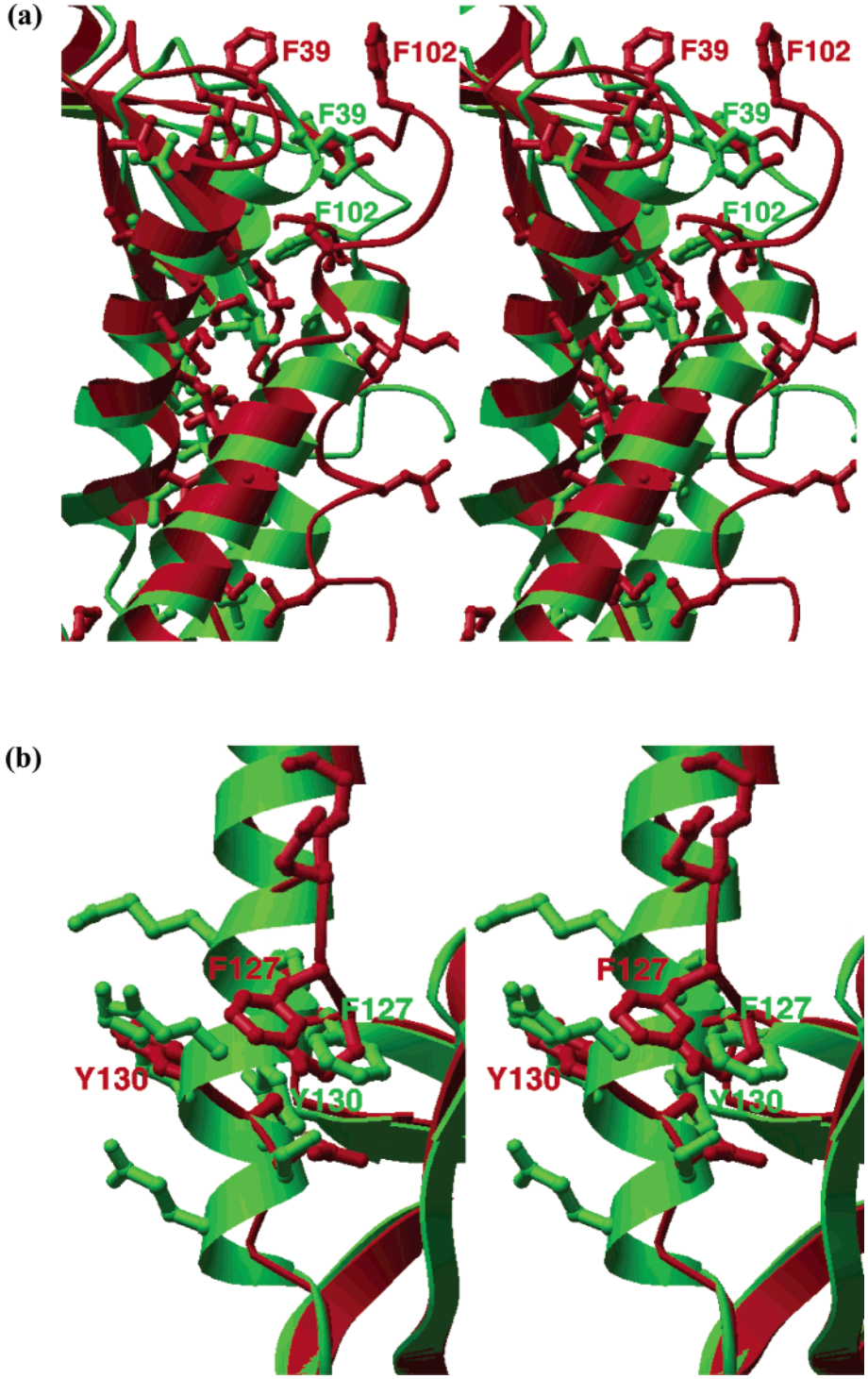

FIGURE 5: Structural comparisons around the large rms regions of two different *Tm*NusAs. (a) Structural comparison of the hydrophobic core in NTD (including residues 1−19 and 98−108). In *Tm*NusA-1, Phe102 is exposed to solvent, while it is hidden in the hydrophobic core in *Tm*NusA-2. The hydrophobic residues are represented by ball-and-stick models. Others are drawn by a ribbon presentation. The green represents *Tm*NusA-2 and the red shows *Tm*NusA-1. Phe39 and Phe102 are labeled. (b) Structural comparison of residues around the connection between NTD and S1 domain. Residues from 125 and 132 are represented by ball-and-stick models. Phe127 and Tyr130 are labeled.

main-chain atoms, and 1.16 Å for all 1597 atoms. However, there are some differences: The S1 domain is connected to a long helical segment (H4) in *Tm*NusA-2 structure, in contrast to an unstructured loop in *Tm*NusA-1 structure (residues 123−132). In the *Tm*NusA-2 structure, Phe127 and Tyr130 interact with the hydrophobic residues from S1 domain. However, in *Tm*NusA-1 structure, these two residues protrude toward the solution (Figure 5b).

These structural differences manifest themselves in the local electrostatic surface potential difference of the NTD between the two structures (Figure 6).

Worbs et al. (*9*) provided a structural point of view in respect to several NusA mutations. There is an interesting mutation in the so-called 449 region, a five amino acid insertion in the *Salmonella typhimurium* NusA between residues 153 and 154 of *E. coli* NusA (*23*). This mutation

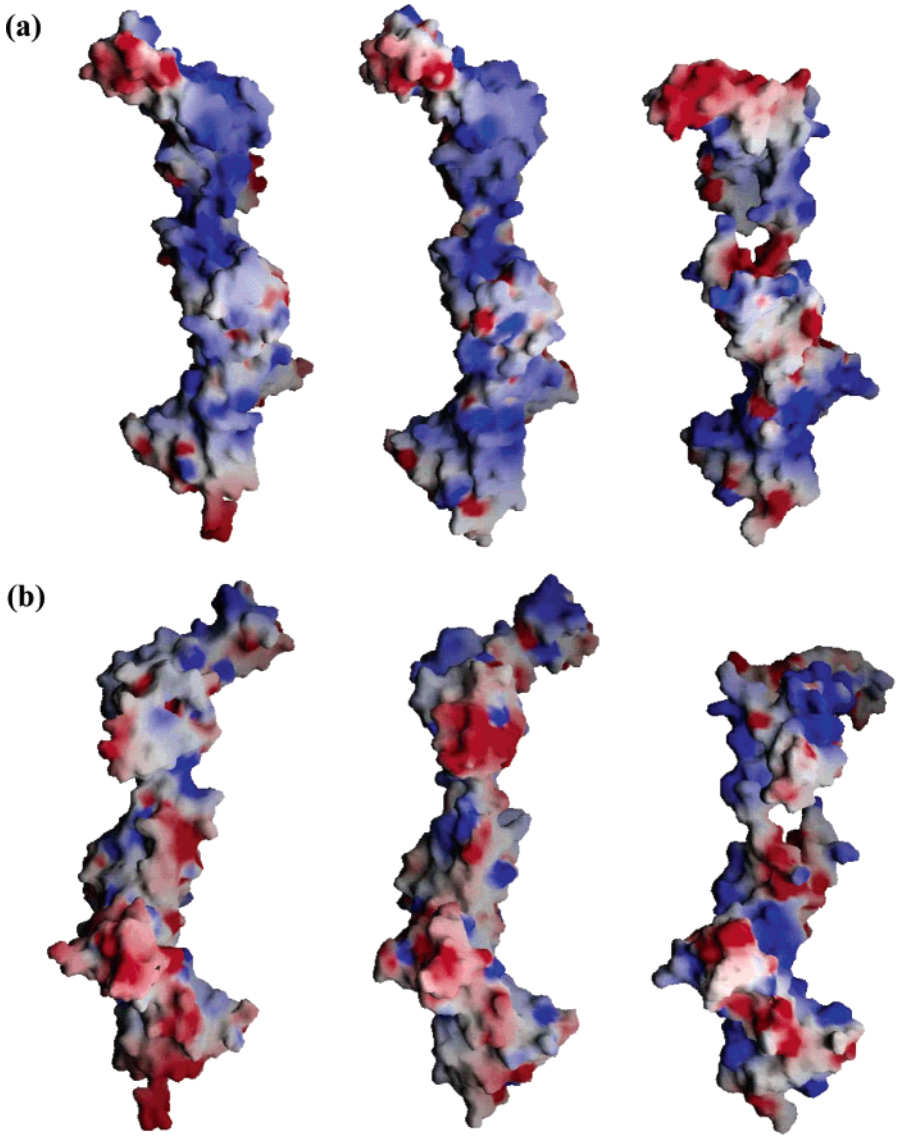

FIGURE 6: Electrostatic surface potential of NusAs. (a) The electrostatic surface potentials of *Tm*NusA-1 (left), *Tm*NusA-2 (middle), and *Mt*NusA (right) are drawn with the program GRASP (*33*) (red, negative; blue, positive; white, uncharged). The globular head domains show a strongly negatively charged surface. (b) The figures were drawn after 180° rotation of the top figures around the vertical axis.

can deform the OB fold of S1 domain as well as disrupt the hydrophobic interaction between H4 and S1 domain in the *Tm*NusA-2 structure. This mutation can affect the dynamics of the linkage between S1 and NTD and thus could affect *Tm*NusA-2's interaction with RNAP.

Structural Comparison between TmNusA-2 and MtNusA. Unlike the elongated shape of *Tm*NusA-2, *Mt*NusA displayed a bent shape between NTD and RBD. Since the *Mt*NusA structure was solved in a different space group, crystal contacts might induce a different conformation from that of *Tm*NusA-2. If KH domains of the two structures are superimposed, the two NTDs are offset by about a 60° rotation. It suggests that NTD and RBD behave like two rigid bodies connected by a linker loop (Figure 4b). However, if each domain of *Mt*NusA is aligned separately with that of *Tm*NusA-2, they are well matched to each other except for the linker regions, indicating that the two structures share a common fold (Figure 4a).

The rms differences between the structures of *Tm*NusA-2 and *Mt*NusA are 2.78 Å for aligned 64 Cα atoms out of 132 NTD Cα atoms and 2.89 Å for aligned 206 Cα atoms out of 212 Cα atoms from S1 and KH domains. These values are 4.77 Å for 63 Cα atoms of the S1 domain (residues from 121 to 128 of a helix from *Mt*NusA being excluded), 0.88 Å for 79 Cα atoms of the first KH domain, and 0.66 Å for 64 Cα atoms of the second KH domain. The major structural differences are observed around NTD (residues 1−132) and S1 domain (Figure 3). *Mt*NusA does not have a globular head domain. Therefore, its NTD is relatively smaller than that of *Tm*NusA-2. Nevertheless, the presence of Glu61, Asp63, and Glu64 at the top of the *Mt*NusA NTD surface results in the same negatively charged electrostatic surface potential as shown in the globular head domain of *Tm*NusA-2 (Figure 6).

The three helices of *Mt*NusA NTD form a helical body domain like that of *Tm*NusA-2. The middle helix (H4 in *Tm*NusA-2) is a six-turn helix in *Mt*NusA and is not connected directly to S1 domain. Instead, a flexible linker loop (residues from 100 to 107) connects the NTD to RBD though it is not visible in the structure (Figure 4b).

The structures of two KH domains are almost the same in both NusA structures (Figure 4a). However, the S1 domain

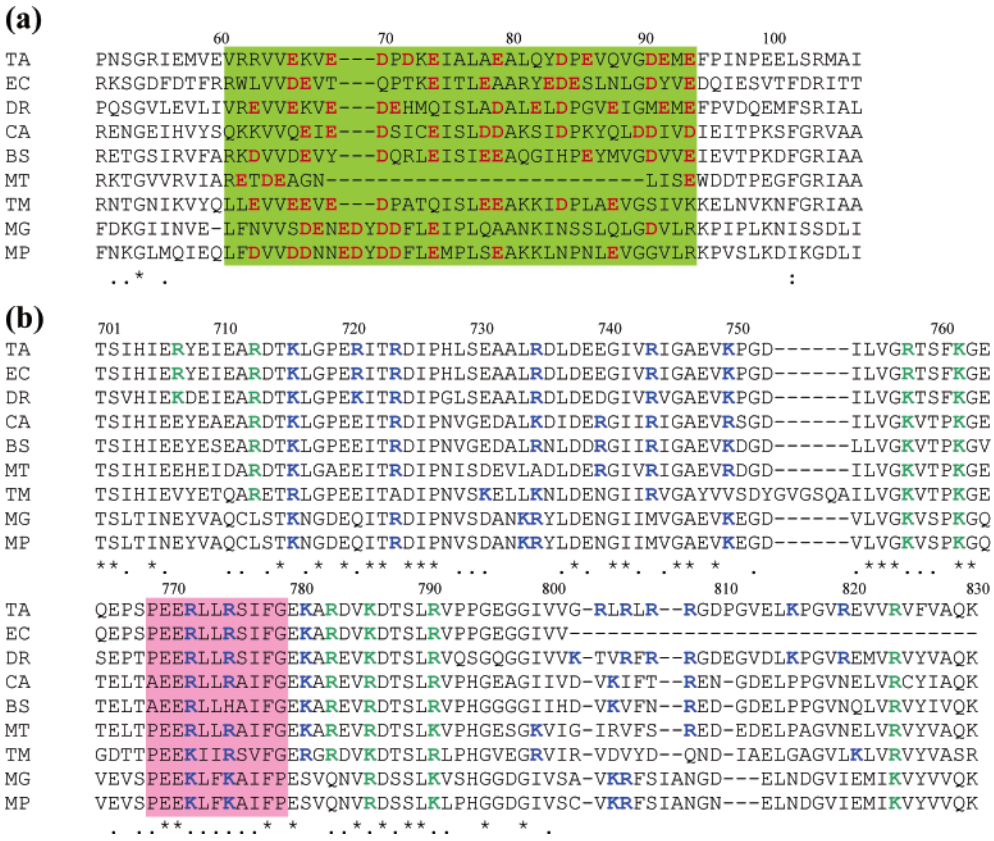

FIGURE 7: Sequence comparisons. (a) Sequence comparison of a globular head region among some NusA homologues from *Thermus aquaticus* (TA), *Escherichia coli* (EC), *Deinococcus radiodurans* (DR), *Clostridium acetobutylicum* (CA), *Bacillus subtilis* (BS), *Mycobacterium tuberculosis* (MT), *Thermotoga maritima* (TM), *Mycoplasma genitalium* (MG), and *Mycoplasma pneumoniae* (MP). The red characters represent negatively charged residues. The green shaded region refers to the globular head domain. The sequence number refers to that of *Thermotoga maritima*. (b) Sequence comparison of a flaplike domain among some RNA polymerases. The flaplike domain is approximately from 704 to 829 based on the crystal structure of *Thermus aquaticus* RNA polymerase. The blue characters represent the positively charged residues exposed to solvent; the green represents residues toward the inside of RNA polymerase. The magenta shaded region refers to the flap-tip helix. The sequence number refers to that of *Thermus aquaticus* RNA polymerase. Dashes in the sequence represent gaps; asterisks are shown below conserved residues, and dots are shown below homologous residues.

of *Mt*NusA has an extra helix between β5 and β6 (*10*) (Figures 3 and 4), an area suggested to be an RNA binding region by Worbs et al. (*9*). The overall electrostatic surface potential of the two structures is slightly different, indicating its general roles in interacting with nonspecific RNA.

Possible Function of NTD. A DALI search (*24*) with NTD was performed to find out a possible molecular function based on structural homologues. The highest resemblance is detected with Lobe II of rad50 ABC-ATPase fragment from *Pyrococcus furiosus* with a *Z* value of 3.0. The rms deviation is 2.3 Å for 50 pairs of aligned Cα atoms of NTD. Although overall topology is similar between the two domains, the NTD of *Tm*NusA has one helix (H3) insertion and does not have a signature motif that binds to a phosphate of ATP as shown in the rad50 ABC-ATPase fragment structure (*25*). Thus, we do not think that the NTD is an ATPase.

The globular head domain of NTD shows a negatively charged surface potential as discussed above. The sequence comparison of the domain from different species shows that the negatively charged residues are positioned around residues 61−94 (Figure 7a). This property may be related to the function of NusA to interact with the positively charged flaplike domain of RNA polymerase. It is reported that a short α-helix at the tip of the flaplike domain covering the RNA exit channel of RNAP contacts a nascent RNA stem−loop structure (hairpin) that inhibits transcription, and this flap-tip helix is required for activity of the regulatory protein NusA (*26*). The sequence alignment of flaplike domains among different species reveals many positively charged residues such as lysine or arginine (Figure 7b). Since the NTD of NusA is known to bind with the core domain of RNA polymerase, the negatively charged globular head domain may be a strong candidate for direct interaction with the flaplike domain of RNA polymerase.

The globular head region also has three hydrophobic patches composed of the following residues; (a) Leu61, Pro85, Leu86, Ala87, and Ile92; (b) Val68, Pro71, Ala72, Leu62, Ala81, Ile75, Ile83, and Val93; and (c) Val64, Val65, Leu77, and Val89. This property may also confer the ability of NusA to interact with hydrophobic patches of RNA polymerase including the flap-tip helix (Figure 7b). The globular head domain is the most nonconserved region in the multiple sequence alignment of NusAs (Figure 7a), suggesting that this region may endow the species specificity of NusAs for their cognate RNA polymerase.

The binding regions of *E. coli* NusA with RNA polymerase have similar properties as those of the σ^{70} factor (*27*). When DALI is run with atomic coordinates of the helical body domain only, one structural homologue is part of α-helices of the σ^{70} factor ($Z = 4.4$). The crystal structure of the core domain of σ^{70} factor revealed that it was

composed of an all-helical structure with a V-shape (*28*). Interestingly, the helices involved in RNA polymerase binding indicate high structural similarity with the helical bundle of *Tm*NusA-2 (helices H12b, H13, and H14 of σ^{70} factor (*28*) approximately matched helices H1, H2, and H4 of *Tm*NusA-2, respectively). The σ^{70} factor (region 2.2) is known to induce RNA polymerase initiation through interaction with a helix bundle (a 48 amino acid coiled-coil from the β' subunit [residues 262−309]) in front of the rudder domain of RNA polymerase (*29*). This implies the importance of the helical secondary structure of the helical body domain of NusA NTD for interaction with RNA polymerase.

CONCLUSIONS

Structural comparison of *Tm*NusA-2 and *Mt*NusA revealed two common structural features of the NTD. First, the top surface of NusA NTDs shows negatively charged electrostatic potential regardless of the structural differences of their globular head domains. Second, both structures form a helical bundle with a hydrophobic core in the helical body domain of the NTD. The structural alignment of the NTD and RBD between *Tm*NusA-2 and *Mt*NusA indicates another feature: the presence of a hinge motion between the middle helix (H4) and S1 domain. As NusA interacts with RNA polymerase and mRNA, the flexibility between domains may facilitate a proper fitting interaction with target molecules. Therefore, we can expect a possible structural transition of both ends of H4 (residues 98−108 and/or 123−132) from helix to loop or vice versa (Figure 4). If this is the case, *Tm*NusA-1 and *Tm*NusA-2 structures may represent two states of the transition. Since RNA polymerase has been observed to undergo significant conformational changes upon switching from σ^{70} factor to NusA (*30*), the flexibility of NusA may enable itself to respond to or induce such changes in RNA polymerase.

ACKNOWLEDGMENT

We thank Dr. Keith Henderson (Advanced Light Source, Lawrence Berkeley National Laboratory) for assistance during data collection. We are also grateful to Barbara Gold for cloning and Bruno A. Martinez for protein expression studies.

REFERENCES

1. Greenblatt, J., and Li, J. (1981) *Cell 24*, 421−428.
2. Greenblatt, J., McLimont, M., and Hanly, S. (1981) *Nature 292*, 215−220.
3. Kingston, R. E., Nierman, W. C., and Chamberlin, M. J. (1981) *J. Biol. Chem. 256*, 2787−2797.
4. Farnham, P. J., Greenblatt, J., and Platt, T. (1982) *Cell 29*, 945−951.
5. Landick, R., and Yanofsky, C. (1987) *J. Mol. Biol. 196*, 363−377.
6. Chan, C. L., and Landick, R. (1993) *J. Mol. Biol. 233*, 25−42.
7. Mah, T. F., Li, J., Davidson, A. R., and Greenblatt, J. (1999) *Mol. Microbiol. 34*, 523−537.
8. Fraser, C. M., et al., and Venter, J. C. (1995) *Science 270*, 397−403.
9. Worbs, M., Bourenkov, G. P., Bartunik, H. D., Huber, R., and Wahl, M. C. (2001) *Mol. Cell 7*, 1177−1189.
10. Gopal, B., Haire, L. F., Gamblin, S. J., Dodson, E. J., Lane, A. N., Papavinasasundaram, K. G., Colston, M. J., and Dodson, G. (2001) *J. Mol. Biol. 314*, 1087−1095.
11. Doublie, S. (1997) *Methods Enzymol. 276*, 523−529
12. Jancarik, J., and Kim, S.-H. (1991) *J. Appl. Crystallogr. 24*, 409−411.
13. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
14. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. D 55*, 849−861.
15. de La Fortelle, E., and Bricogne, G. (1997) *Methods Enzymol. 276*, 472−494.
16. Abrahams, J. P., and Leslie, A. G. W. (1996) *Acta Crystallogr. D 52*, 30−42.
17. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A 47*, 110−119.
18. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D 54*, 905−921.
19. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *Appl. Crystallogr. 26*, 283−291.
20. Luzzati, V. (1952) *Acta Crystallogr. 5*, 802−810.
21. Subramanian A. R. (1983) *Mol. Biol. 28*, 101−142.
22. Gusarov, I., and Nudler, E. (2001) *Cell 107*, 437−449.
23. Craven, M. G., Granston, A. E., Schauer, A. T., Zheng, C., Gray, T. A., and Friedman, D. I. (1994) *J. Bacteriol. 176*, 1394−1404.
24. Holm, L., and Sander, C. (1997) *Nucleic Acids Res. 25*, 231−234.
25. Hopfner, K. P., Karcher, A., Shin, D. S., Craig, L., Arthur, L. M., Carney, J. P., and Tainer, J. A. (2000) *Cell 101*, 789−800.
26. Toulokhonov, I., Artsimovitch, I., and Landick, R. (2001) *Science 292*, 730−733.
27. Traviglia, S. L., Datwyler, S. A., Yan, D., Ishihama, A., and Meares, C. F. (1999) *Biochemistry 38*, 15774−15778.
28. Malhotra, A., Severinova, E., and Darst, S. A. (1996) *Cell 87*, 127−136.
29. Young, B. A., Anthony, L. C., Gruber, T. M., Arthur, T. M., Heyduk, E., Lu, C. Z., Sharp, M. M., Heyduk, T., Burgess, R. R., and Gross, C. A. (2001) *Cell 105*, 935−944.
30. Gill, S. C., Weitzel, S. E., and von Hippel, P. H. (1991) *J. Mol. Biol. 220*, 307−324.
31. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
32. Carson, M. (1991) *J. Appl. Crystallogr. 24*, 958−961.
33. Nicholls, A. (1991) *Proteins: Struct., Funct., Genet. 11*, 281−296.

BI035118H